

**Slide 1 of 13**

Title Slide: Considerations for Analysis of Postsecondary Education Data

**Slide 2 of 13**

This module describes analytic considerations that should be kept in mind when using the NPSAS, BPS, or B&B restricted-use, micro-level data. This module describes the analysis considerations for: using the derived files, using the source files, considering the “Applies To” fields and reviewing frequencies, generalizing beyond defined sampling strata, noting study changes across administrations, considering trend analyses, using postsecondary education transcripts, and understanding weight component files.

Throughout this module, underlined blue screen text indicates a link to additional resources.

**Slide 3 of 13**

As noted earlier, NCES strongly advises users of micro-level data to use the derived file for analysis whenever possible.

Users of longitudinal data-sets will notice that their data products contain multiple files, including multiple derived files. This includes the derived file from the base year data collection, as well as cumulative derived files for subsequent follow-ups. That is, a first-follow-up data product would contain a base year derived file as well as a first-follow-up derived file that incorporates both data derived from the current collection and variables from the base-year derived file. The name and description of the derived files on the data product will provide you information about what each derived files contains, as well as indicate the final derived file.

File names typically contain information about the administration year (e.g., N4 is the code for the NPSAS:04 derived file, with F6 and F9 corresponding to BPS follow-ups in 2006 and 2009, respectively). Since the derived files are cumulative, only the file listed first within the README file should be of concern to researchers.

**Slide 4 of 13**

Users who go beyond the derived file and choose to use source files should use caution. These variables are used by NCES and its data collection contractors to prepare the derived file for analysts, and are typically included on the data product without edits to variable names, labels, descriptions, or data values.

Additionally, users are cautioned to carefully review the syntax for, and the results of, MERGE commands across source files. The ECB software, if used, will only merge records containing student-level data. Other types of source file data merges, such as student-school, must be done manually. Even when using the ECB software’s MERGE syntax, analysts are cautioned to compare the number of resulting records in the MERGED data set to the number of records anticipated at the outset of the process.

**Slide 5 of 13**

It is important to review all information provided about a derived variable, including descriptive statistics.

All NCES derived variables include an “Applies To” note that specify the subpopulation of study members for whom a valid value should be expected. Particularly when the variable in question represents a dollar value – such as student aid or undergraduate debt – analysts are advised to think carefully about the meaning of a zero value and how that may nuance the interpretation of any statistic. For example, the average of a “total student aid” variable excluding zeros indicates the average amount of aid among aided students. Including zeros would create an average across all students, irrespective of whether they were aid recipients.

Finally, analysts using the restricted-use CD are reminded to check for unweighted frequencies, both to identify special data values that should be excluded from statistics (-3s and -9s, for example), as well as to ensure that a sufficient number of cases are present for analysis.

**Slide 6 of 13**

The NPSAS family of studies has been designed to generate nationally, and sometimes state, representative estimates for various types of institutions and subsets of students. In particular, NPSAS is designed to be representative of undergraduate and graduate students at each major type of institution, BPS of first-time beginning college students, and B&B of baccalaureate graduates. Although other student and institution subpopulations of interest can be found in the data, they were not typically explicitly stratified or otherwise sampled for. Analysts are reminded to consult each study’s methodology report to understand the sampling strata used in its development. Techniques such as implicit stratification, where cases are ordered on sampling lists by key variables, can help ensure coverage of important student or institution characteristics and improve the quality of the resulting estimates.

Particularly when estimates may be based on relatively low sample sizes, analysts are encouraged to present their results with standard errors or confidence intervals to provide readers information about the precision of a given estimate. Similarly, analysts are encouraged to use good judgment when describing findings to avoid over-generalizing specific conclusions.

No study in the NPSAS family is designed to produce accurate institution-level estimates. As a result, never use any sample survey product to make statements about specific schools, colleges, or universities.

**Slide 7 of 13**

The NPSAS family of studies is dynamic, changing in response to changes in federal financial aid policy, evolving research questions, technologies, and logistical or operational constraints.

Between NPSAS:2000 and today, analysts should be cognizant of several changes to NPSAS and its related longitudinal studies.

Notable changes include those regarding institutional and student eligibility, disproportionate sampling rates, and Puerto Rican institutions.

The most significant change in institutional eligibility occurred in 2000, which limited institutions to those participating in Title IV Federal Student Aid Programs. The most significant change in student eligibility occurred in 2012 when NCES clarified that institutions should consider eligible students enrolled entirely in remedial coursework so long as their program of study was otherwise Title IV eligible.

Specific policy needs in a given year may motivate student oversamples, such as a focus on SMART grant recipients in NPSAS:08 or sub-baccalaureate students in NPSAS:12. Some studies always oversample to increase the likelihood of yielding key student populations, such as potential teachers in Baccalaureate and Beyond.

In 2012, Puerto Rican institutions were not sampled in NPSAS. For consistent estimates across time, analysts should use the COMPTO87, spelled C-O-M-P-T-O-8-7, variable in prior NPSAS studies to ensure only institutions in the 50 States and the District of Columbia are included.

Each study's methodology report documents how the present study differs from its predecessors, if at all, and what steps analysts can take to improve the comparability of time series estimates.

**Slide 8 of 13**

Two other cautions are important to note for analysts conducting trend analyses.

First, analysts who seek to compare dollar amounts over time, such as tuition costs or loan amounts, are advised to consider the effects of inflation on those estimates. The Bureau of Labor Statistics provides tools to help users adjust their data for inflation, including on-line calculators and detailed Consumer Price Index tables.

Second, users are reminded that federal financial aid policy has shifted over time, including the introduction and retirement of various loan and grant programs, as well as changes in those programs' statutory limits. Analysts are cautioned to consider the effect a constantly changing aid context may have for the interpretation of estimates and the development of findings.

### Slide 9 of 13

Many analyses focus on specific subpopulations, such as students beginning at a certain type of institution or students receiving a particular type of aid.

When analyzing data based on complex samples, each case contains important information needed to properly calculate variance estimates. Rather than deleting cases from your dataset that are not part of your subpopulation of interest, or using common sub-setting commands such as “if” to limit cases entering your analysis, be sure to use the subpopulation selection command that is part of your software package. In Stata, this command is “subpop,” while in SAS the command is “domain.” For additional information, consult your software package’s technical documentation in conjunction with the resource document that can help you use sub-setting commands in SAS and Stata to properly analyze postsecondary education datasets.

### Slide 10 of 13

Transcript data accompany both the 2007-08 iteration of Baccalaureate and Beyond and the 2012-2017 Beginning Postsecondary Students Longitudinal Study. Transcripts were not collected for the 2015-16 B&B cohort.

As part of B&B, NCES requested students’ bachelor’s degree transcript. In contrast, NCES asked for transcripts from all known institutions attended by BPS students.

Analysts using the BPS transcripts should be aware that they should select one of two weights when analyzing transcript data. If transcripts are being analyzed without any accompanying student interview data, the appropriate weight is WTC. When transcript data is analyzed in combination with any other BPS data element, the appropriate weight is WTD.

### Slide 11 of 13

Finally, users of restricted-use micro-level data are reminded that they have access both to each study member’s final analysis weight – that is, the weight that includes the probability of institutional and student selection and nonresponse and post-stratification adjustment – but also each of the weight components.

As noted in each study’s methodology report, each weight component is multiplied together to create the final analysis weight. While this final weight is sufficient for most purposes, analysts who wish to evaluate multi-level models or calculate their own estimates of nonresponse bias may require each weight component separately. Each data product’s WEIGHT file contains weight components, labeled WT1, WT2, WT3, and so forth, as well as the final analysis weight.

More details on how each weight component was derived can be found in each study’s methodology report.

**Slide 12 of 13**

This module has described the analytic considerations that should be kept in mind when using the NPSAS, BPS, or B&B restricted-use, micro-level data. Specifically, this module described the analysis considerations for: using the derived files, using the source files, considering the “Applies To” fields and reviewing frequencies, generalizing beyond defined sampling strata, noting study changes across administrations, considering Time Series Analyses, using postsecondary education transcripts, and understanding weight component files.

**Slide 13 of 13**

In addition, it has provided resources that can be accessed through the DLDT system, the NCES website, and /or the Internet.

You have now completed this series of modules, click the exit button to return to the landing page.